(54) Title: METHOD AND SYSTEM FOR IDENTIFYING DOCUMENTS GENERATED BY AN UNAUTHORIZED SOFTWARE COPY
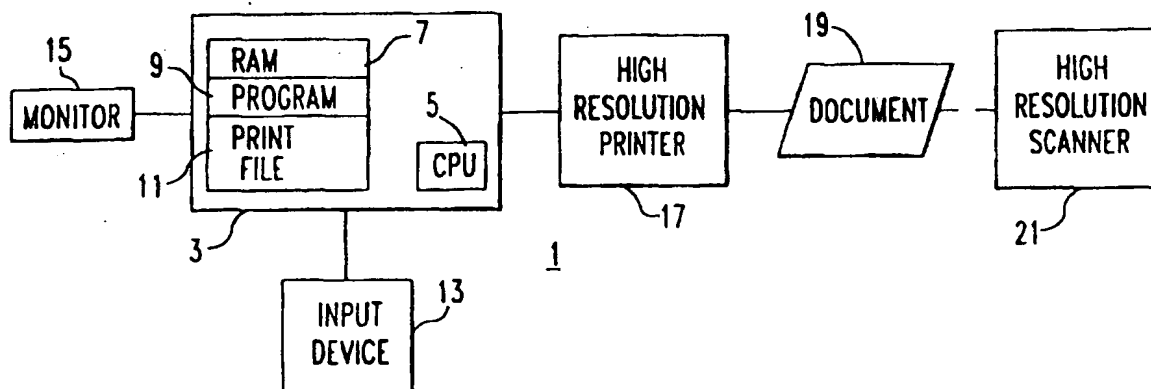
(57) Abstract

Each copy of software (9) is assigned a unique identifying code pattern (27) which is printed on all documents (19) produced with that software (9) by a high resolution printer (17). The unique identifying code pattern (27) is a plurality of spaced apart marks having a size no greater than about 300 dpi, and is therefore, at best, barely noticeable to the human observer. The "invisible signature" is also reproduced on documents (19) made by unauthorized copies of software (9) which can therefore be traced. Preferably, the unique identifying code (27) is replicated multiple times (27_1-27_5) over the document (19) using an error correcting code to assure that at least one replication will be clear of matter selected for printing by the software. A high resolution scannner (21) extracts and identifies the code patterns (27_1-27_5) printed on the document (19). In systems where the software (9) generates (39) a print file for the high resolution printer (17), print commands for the pattern replications (27_1-27_5) are interspersed (41) with the other print commands making identification and removal of the commands very difficult and not worth the effort since the "invisible signature" does not prevent copying of the software (9) or noticeably detract from the appearance of the finished document (19).

- 1 -

METHOD AND SYSTEM FOR IDENTIFYING DOCUMENTS GENERATED BY AN UNAUTHORIZED SOFTWARE COPY

BACKGROUND OF THE INVENTION

Field of the Invention

This invention relates to a method and system for software protection through identification of unauthorized copying of software. More particularly, it relates to tracking down unauthorized copies of software through identification of documents generated by the unauthorized software copy on which a unique identifying code is imprinted by a high resolution printer.

Background Information

Software piracy is one of the biggest problems facing the software industry. It is estimated that only about 40% (by value) of the software in the United States is legal. That is only about 40% of the software has been purchased. The remainder is copied from legal software. As low as this may sound, it is a higher percentage of any other country. In Britain the estimate is that only about 20% of the software is legal. The Far East and Russia are known as one diskette countries, meaning that almost no legal software exists. According to a recent estimate, the cost to the software industry of piracy is $10 - 12 Billion a year.

Present techniques for protecting software are principally directed toward making it more difficult to copy the software. However, there is considerable consumer resistance to this approach, and computer hackers take great pride in meeting the challenge of defeating the lock. Other approaches such as placing a "time bomb" in the program which is activated if a license fee is not timely paid are also not

received favorably by purchasers and can lead to possible liability for destroying a user's data.

The software industry has recently tried to stem the tide of piracy through legal remedies. However, this is a costly and time-consuming approach, which generally requires access to the host computer, and could backfire if suspicions turn out to be unfounded. Generally, legal redress is only suitable for large scale piracy.

There is a need, therefore, for an improved technique for protecting software from piracy.

There is a need for such a technique which does not require access to the host computer on which the software is run.

There is also a need for such a technique which is sufficiently secure that attempts to defeat it are discouraged.

There is an additional need for such a technique which allows identification of unauthorized software copies without noticeably affecting the performance of the software or even the ability to copy it, again, so that the incentive to by-pass the technique is minimized.

SUMMARY OF THE INVENTION

These needs and others are satisfied by the invention which is directed to uniquely identifying each authorized software copy by an essentially invisible signature applied to the documents generated by that software copy. In particular, the signature which is basically invisible to the unaided human eye, is a unique identifying code pattern which is imprinted on the document. The invention is designed to be used with a high resolution printer, which for the purposes of this application is defined as a printer having a resolution at least as great as about 300 dpi (dots per inch). The unique identifying code pattern is made up of non-contiguous marks, preferably single dots, generated by such a high resolution printer, however, when printers with a higher resolution are used, each mark may comprise multiple adjacent dots as long as they collectively constitute a mark having a size no greater than about 300 dpi.

The unique identifying code pattern is automatically applied to each document made by the software. Unauthorized copies of the software will also generate documents with the same unique identifying code pattern thereon. Printers commonly used today have at least 300 dpi resolution and some have resolutions of 600 or 1000 dpi. Even at 300 dpi, the non-contiguous marks constituting the invisible

signature are barely visible and go unnoticed by the casual observer. Documents produced by software incorporating the invention, can be scanned with a high resolution scanner which extracts the unique identifying code pattern for identification. Documents produced by a party not having a licensed copy of the software thus can be traced.

The unique identifying code can be placed on the document in the an area in which the software application will typically not print. Preferably, however, multiple replications of the unique identifying code pattern are spaced across the document so that the likelihood is increased that the "invisible signature" will not be obscured by other matter printed on the document. Also, the signature should be imprinted using an error correcting code so that imperfections in the paper and "smears" of the printer will not affect it.

The invention embraces both a method and a system for printing a unique identifying code pattern on the document using a high resolution printer. In a broader respect, the invention is directed to a method of identifying a medium used to generate an image by employing the medium to generate an image incorporating the unique identifying code pattern, and scanning the image to extract and identify that pattern.

While the invention is only applicable to a medium which generates an image, such as software which generates a document, it has many advantages over present techniques. It does not require access to the equipment which generated the document or image. It does not prevent full use, or even copying, of the software or medium. Preferably, in software which generates a print file containing commands for the printer, print commands for the unique identifying code pattern are interspersed in the print file which makes it very arduous and time-consuming to attempt to extract those commands. The incentive is clearly to discourage attempts to defeat the system which leaves the user of an unauthorized copy open to detection.

BRIEF DESCRIPTION OF THE DRAWINGS

A full understanding of the invention can be gained from the following description of the preferred embodiments when read in conjunction with the accompanying drawings in which:

Figure 1 is a schematic diagram in block form of a system for implementing the invention.

Figure 2 is an illustration of a document generated by the system of Figure 1 and incorporating a unique identifying code in accordance with the invention and with one replication of the code magnified.

Figure 3 is an example of a code which can be imprinted on the document of Figure 2 by the system of Figure 1 in accordance with the invention.

Figure 4 is a flow chart of a routine used to generate a print file used by the system of Figure 1 in accordance with the invention.

DESCRIPTION OF THE PREFERRED EMBODIMENT

The invention will be described as applied to protecting software, for which it is particularly suitable. The software is utilized in the computerized document production system 1 illustrated in Figure 1. This system includes a digital computer 3 which includes a central processing unit (CPU) 5 and random access memory (RAM) 7 which stores the software program 9 for use by the CPU 5, and a print file 11. The CPU 5 implements the program 9 in accordance with instructions entered by the user through an input device 13 which can include one or more of a keyboard, a mouse, a track ball or a touch screen, for instance. The program 9 to which the invention applies has the capability of generating a printed document. This can be in addition to a display generated on a monitor 15. In fact, it is not necessary that the program always generate a document, but only that it is capable of generating a printed document. A high resolution printer 17 generates the document 19 in accordance with the information provided to it by the computer 3. Typically, the computer generates the print file 11 which contains detailed commands to the printer of the matter to be printed on the document 19. This matter could be text, numeric, alphanumeric, a pictorial image or any other matter. By high resolution printer, is meant that the printer has resolution of at least 300 dots per inch (dpi). Currently available laser printers have at least this resolution and are also available with resolutions of 600 dpi and 1000 dpi. Also, it is not necessary that the computer generate a print file 11 for transmittal to the high resolution printer. Where the printer has its own software for generating the print instructions, the computer need only send to the printer the information to be printed.

As will be discussed in more detail, embedded in the printed document is a unique identifying code pattern assigned to the particular authorized copy of the software. The identity of the authorized owner can be recorded at the time of purchase

of the software, or incentives can be provided to the purchaser to register the software after purchase such as by offering updates to the program only to those who register their name and address. The unique identifying code pattern is in essence a serial number or identifier for the authorized copy which becomes an invisible signature on the printed document 19.

While this "invisible signature" is not readily apparent to the casual observer of the document, and indeed with very high resolution printers may be imperceptible, the unique identifying code pattern can be extracted from the document 19 by a high resolution scanner 21. Such a scanner must have a resolution of at least as great as that of the printer. Scanners with such resolution are readily available, and in fact there are commercial models with resolutions as high as 2400 dpi.

Figure 2 illustrates a document 19 generated by the system of Figure 1 in accordance with the invention. The exemplary document 19 is a printed document with lines of text 23 surrounded by clear margins 25. The unique identifying code pattern 27 is printed on the document simultaneously with the text 23. A single pattern 27_1 can be printed at a location where it will be clear of the text 23. However, where the program prints selected matter other than printed text, such as, for example an image, there is no guarantee that any area on the document will always be free of selected matter 23. Therefore, it is preferred that multiple replications 27_1 - 27_5 be printed at spaced apart locations across the face of the document 19. The number of replications shown in Figure 2 and their locations are exemplary only. It is within the contemplation of the invention that a wide variety of the number of replications and their locations could be utilized. The selection of the number of replications and their locations could also be a function of the particular application program, with the selection being made to maximize the number of replications likely to appear in clear space on the document. As can be seen in Figure 2, some of the replications of the unique identifying code pattern such as 27_1 and 27_5 appear in the margins. Others such as 27_2 appear in the field of the printed text but between lines so that they are not obscured by the text. Still others such as 27_3 and 27_4 are at least partially obscured by the matter 23 selected by the program for the image to be generated on the document.

The replication 27_1 of the unique identifying pattern is shown magnified in Figure 2 to illustrate that it is composed of a pattern of non-contiguous marks 29 each having a resolution at least as great as 300 dpi. Where the printer 17 is a 300 dpi printer, each of the marks 29 is a single dot. If a higher resolution printer is used, for instance, a 1000 dpi printer, each of the marks 29 may be formed by adjacent dots, just so that collectively they have a resolution at least as great as 300 dpi. Thus, for instance, for a 1000 dpi printer each of the marks 29 could be 1, 2, or 3 dots. Preferably, however, single dots with a higher resolution printer are used, such as, for example single dots at 1000 dpi, so that the unique identifying code pattern 27 is imperceptible to the human eye, but can be read by a scanner 21 with at least a 1000 dpi resolution.

While any code pattern could be utilized, the exemplary unique identifying code 27 is a series of binary coded decimal numbers. For example, as shown in Figure 3 such binary coded decimal numbers can be represented in columns having rows indicating by dots the powers of 2 which make up the number represented by that column. Thus, for the representative number 607958, the six is represented by the left most column with dots 29 in the two row and four row. An odd parity check sum column, CK, to the right of the least significant number provides a measure of reliability in reading the code. Preferably, a error correcting code can be used.

An example of a simple error correcting code is shown in Figure 3. In addition to the odd parity check sum column, CK, an odd parity check sum row, RW, can also be provided in the unique identifying code pattern. Assume for purposes of illustration that the detected pattern has an error in that a dot $29'$ is erroneously detected in the first power of two row for the right hand column so that the detected least significant digit is 9 rather than 8. In this instance, the first power of two row odd parity check computed by the scanner from the detected pattern will not agree with that printed in the detected pattern. In addition, the parity check for the right-hand column will also not check in that a dot is found in the right-hand column odd parity check while the detected dot pattern would predict a 0 in this column. Hence, the row and column parity bits which intersect at the first power of two location in the least significant bit indicate that this dot $29'$ is an error and should not be there. Thus, the scanner can correct the detected code. If multiple dots are in error, such as could

occur if the unique identifying code pattern overlaps printed matter selected by the program, it may not be possible to correct the code. However, with the multiple replications of the code, similarities between detected patterns can be used to verify the code. In its simplest form, if two patterns are the same, the code is verified.

5   If the replications 27 of the unique identifying code are always placed in the same location on the document, the scanner 21 can be programmed to interrogate those locations. However, if the locations of the unique identifying code 27 are random, or vary with different manufacturers or programs, each replication of the code can be preceeded by a marker such as the symbol 31 shown in Figure 3 to indicate to

10   the scanner the beginning of a replication of the unique identifying code pattern. Of course, any other symbol can be used as the marker to indicate the location of the unique identifying code patterns.

The unique identifying code pattern 27 with the marker 31, if provided, is placed in the print file by a routine 33 such as that illustrated by the flow chart of

15   Figure 4. The routine reads the input of the "invisible signature" at 35. For instance, in the pattern illustrated in Figure 3 the program would read the number "607,958". The signature is then encoded, for instance in binary coded decimal, at 37. A print function for the binary coded decimal pattern plus the marker (such as 31 shown in Figure 3) is then generated at 39. This print function contains commands, for instance

20   in the postscript language where the print file is in postscript, telling the printer how to print the desired pattern and marker. Each time the print file is generated by the software for printing a document, the print function generated at 39 is inserted into the print file. In the exemplary routine, the unique identifying code pattern is inserted at random into the print file. While this command is inserted at random into the print file

25   it could be printed at a fixed location in the document depending upon the command. The printer when reading the print file sorts out the various commands and arranges them in order for printing on the document. In the example, multiple replications of the unique identifying code pattern 27 are inserted in the document, hence, the routine loops back at 43 until all replications have been inserted into the print file. The routine

30   is then exited at 45.

The unique identifying code pattern or "invisible signature" can be added to any document produced by a specific software. This will make it possible to detect the particular software copy that produced the document by checking the hard copy

itself. As access to a document should be much easier than to the computer on which the software produced it resides, it will become easier to identify software pirates. If a producer of software suspects that someone has an unauthorized copy of its software, it could send an inquiry to the user which would require a response using the software,

5   such as, for example a letter. The letter could then be scanned with a high resolution scanner to extract the unique identifying code pattern which could then be checked against the registered owner of that copy of the software. If the sender of the document is not the registered owner, then an investigation could be made into how the user gained access to the program copy used. The invisible signature will be

10   reproduced by copying of the document using a copier of reasonable quality. In fact, depending upon the quality of the copier, even copies of the copies should retain the signature. After several generations of copies on a good printer or even the first copy of a printer with poor resolution, the "invisible signature" will probably be degraded to the point that it could not be extracted by scanner, but then again, the overall quality

15   of the document will also be seriously degraded.

While specific embodiments of the invention have been described in detail, it will be appreciated by those skilled in the art that various modifications and alternatives to those details could be developed in light of the overall teachings of the disclosure. Accordingly, the particular arrangements disclosed are meant to be

20   illustrative only and not limiting as to the scope of invertion which is to be given the full breadth of the claims appended and any and all equivalents thereof.

What is Claimed is:

         1.   A method of identifying a software copy used to generate documents (19) comprising the steps of:

        embedding (41) in said software a unique identifying code pattern (27) to be printed in addition to matter selected for printing by said software;

        generating a printed document (19) containing said unique identifying code pattern (27) and said matter (23) selected for printing by said software, said unique identifying code pattern (27) comprising a plurality of non-contiguous marks, each having a size no greater than about 300 dpi; and

        scanning said document (19) to extract and identify said unique identifying code pattern (27).

        2.   The method of Claim 1 wherein said step of generating a printed document (19) comprises generating said document (19) with multiple replications $(27_1 - 27_5)$ of said unique identifying code pattern (27) spaced apart over said document (19).

        3.   The method of Claim 2 wherein said step of scanning said document (19) comprises extracting said multiple replications $(27_1 - 27_5)$ of said unique identifying code pattern.

        4.   The method of Claim 3 wherein said scanning step comprises identifying said unique identifying code pattern replications $(27_1 - 27_5)$ based upon predetermined similarity of said multiple replications $(27_1 - 27_5)$ of said unique identifying code pattern (27).

5.   The method of Claim 4 wherein said predetermined similarity is established when at least two of said replications $(27_1 - 27_5)$ of said unique identifying pattern code (27) are the same.

6.   The method of Claim 1 wherein said step of generating the printed document (19) comprises printing said unique identifying code pattern (27) as a pattern of single non-contiguous dots each having a size no greater than 300 dpi.

7.   The method of Claim 1 wherein said step of generating said document (19) comprises printing said document with a printer (17) having a resolution greater than about 300 dpi with each mark (27) comprising contiguous dots which collectively have a size no greater than about 300 dpi.

8.   The method of Claim 1 wherein said unique identifying code pattern (27) embedded in said software comprises a plurality of binary coded decimal numbers.

9.   The method of Claim 1 wherein said software generates a print file containing commands for printing and said step of embedding comprises embedding commands in said print file for generating said unique identifying code pattern (27), and wherein generating said document (19) comprises implementing said commands to generate said unique identifying code pattern (27) with a printer (17) having a resolution at least as great as about 300 dpi.

10.   The method of Claim 9 wherein multiple replications $(27_1 - 27_5)$ of said unique identifying code (27) are printed by said printer (17) in spaced relation on said document (19).

11.   A system (1) for identifying a software copy used to generate a document (19) comprising:

a software copy (9) having embedded therein instructions (33) for creating a unique identifying code (27) comprising a plurality of non-contiguous marks each having a size no greater than about 300 dpi;

a digital computer (3) for running said software (9);

user input means (13) for controlling running of said software (9) in said digital computer (3) to select matter (23) to be printed in addition to said unique identifying code pattern (27);

a high resolution printer (17) having a resolution of at least 300 dpi for printing a document (19) containing said matter (23) selected to be printed and said unique identifying code pattern (27); and

scanning means (21) for extracting from said document (9) said unique identifying code pattern (27).

12.   The system (1) of Claim 11 wherein said high resolution printer (17) prints multiple replications ($27_1 - 27_5$) of said unique identifying code pattern (27) spaced apart on said document (19), and wherein said scan means (21) extracts said multiple replications ($27_1 - 27_5$) of said unique identifying code pattern (27) from said document (19).

13.   The system (1) of Claim 11 wherein said software (9) has a print file containing print commands for printing said matter (23) selected for printing and for printing said unique identifying code pattern (27), and wherein said high resolution printer (21) executes said print commands to generate said document (19) with said matter (23) selected and said unique identifying code pattern (27).

14.   The system (1) of Claim 11 wherein said unique identifying code pattern (27) printed by such high resolution printer (17) is a plurality of binary coded decimal numerals.

15.   A method of identifying a medium (9) used to generate an image (23), comprising the steps of:

employing said medium (9) to generate said image (23) incorporating a unique identifying code pattern (27) of non-contiguous marks, with each of said marks having a size no greater than about 300 dpi; and

scanning said image (23) to extract and identify said unique identifying code pattern (27).
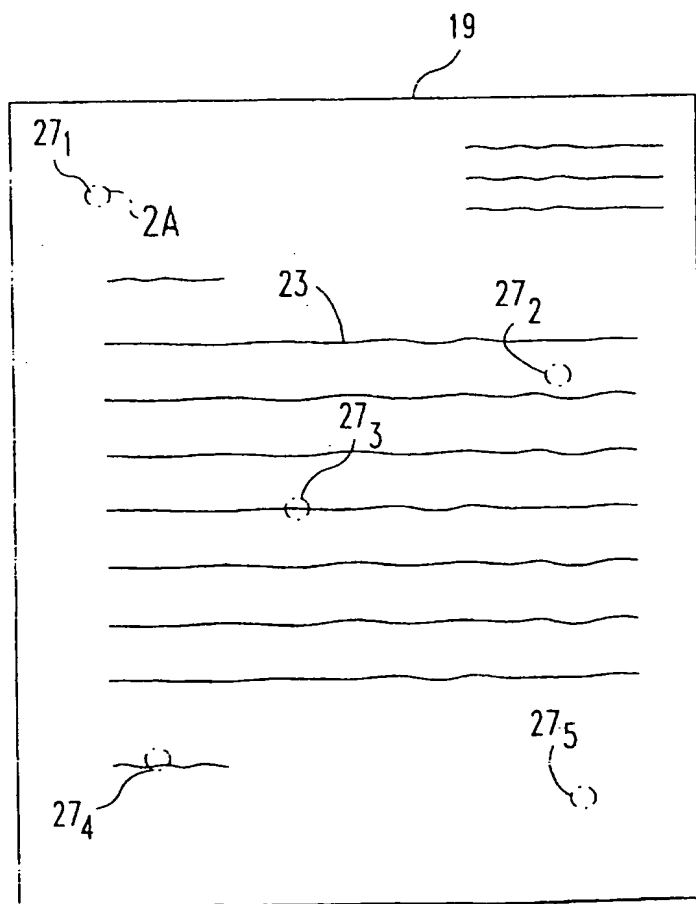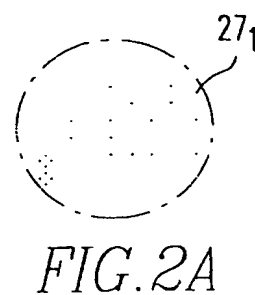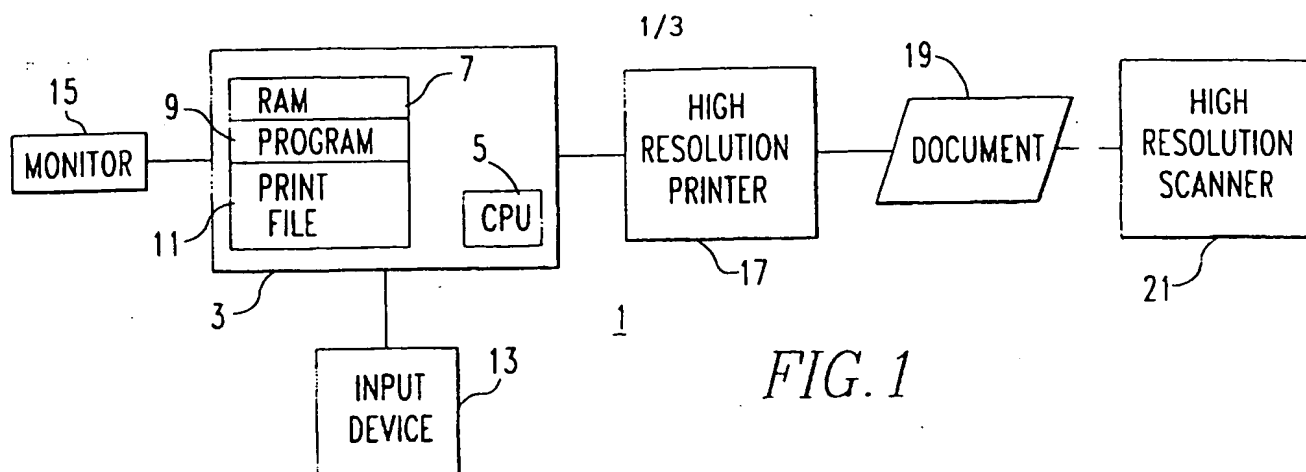
16.   The method of Claim 15 wherein said medium (9) is employed to generate said image (23) incorporating multiple replications ($27_1 - 27_5$) of said unique identifying code pattern spaced (27) apart in said image, and said step of scanning includes extracting and identifying said multiple replications ($27_1 - 27_5$) of said unique identifying code pattern (27).

17.   The method of Claim 16 wherein said step of scanning includes extracting said multiple replications ($27_1 - 27_5$) of said unique identifying code pattern

(27) and identifying said unique identifying code pattern (27) in response to a predetermined similarity of the multiple replications $(27_1 - 27_5)$ of said unique identifying code pattern (27) extracted.

18.   The method of Claim 15 wherein said unique identifying code pattern (27) comprises a plurality of binary coded decimal numerals.

5

1/3

FIG.1

FIG.2A

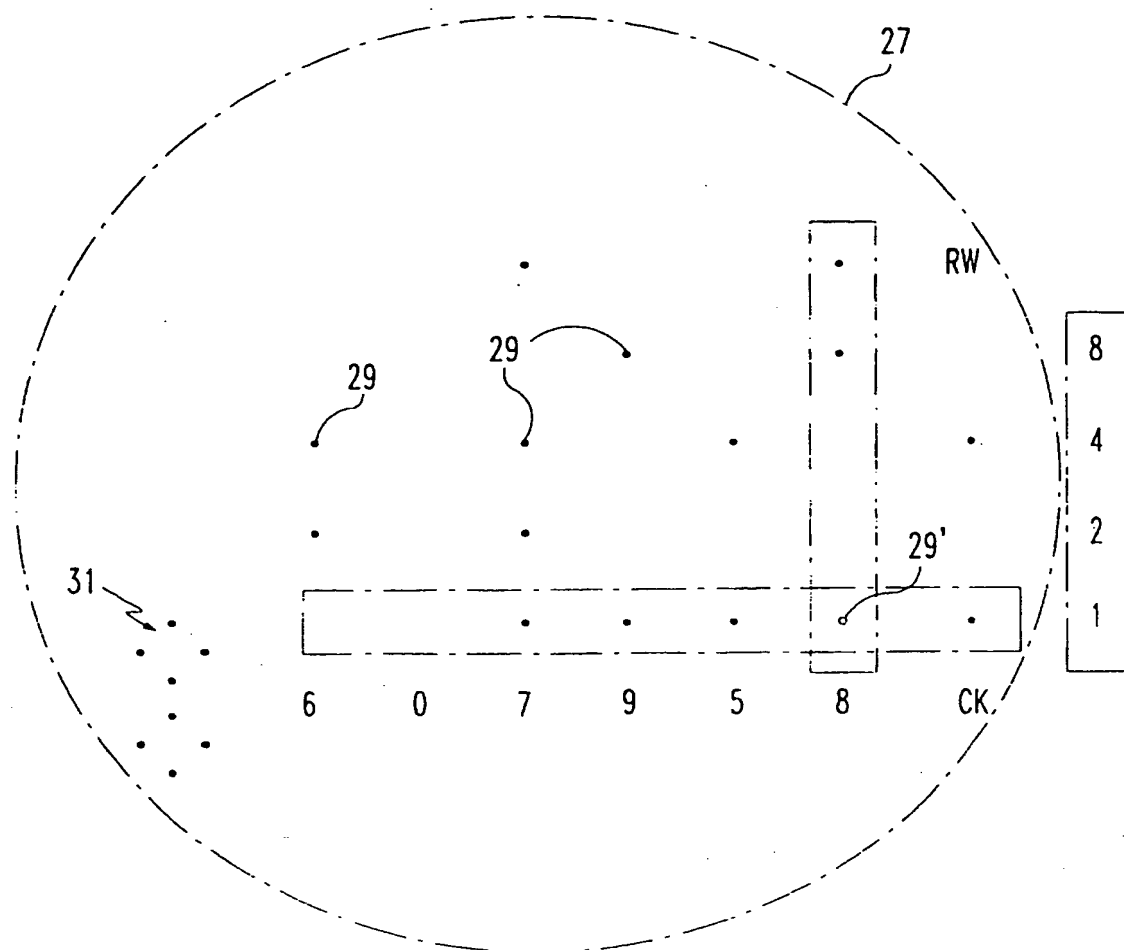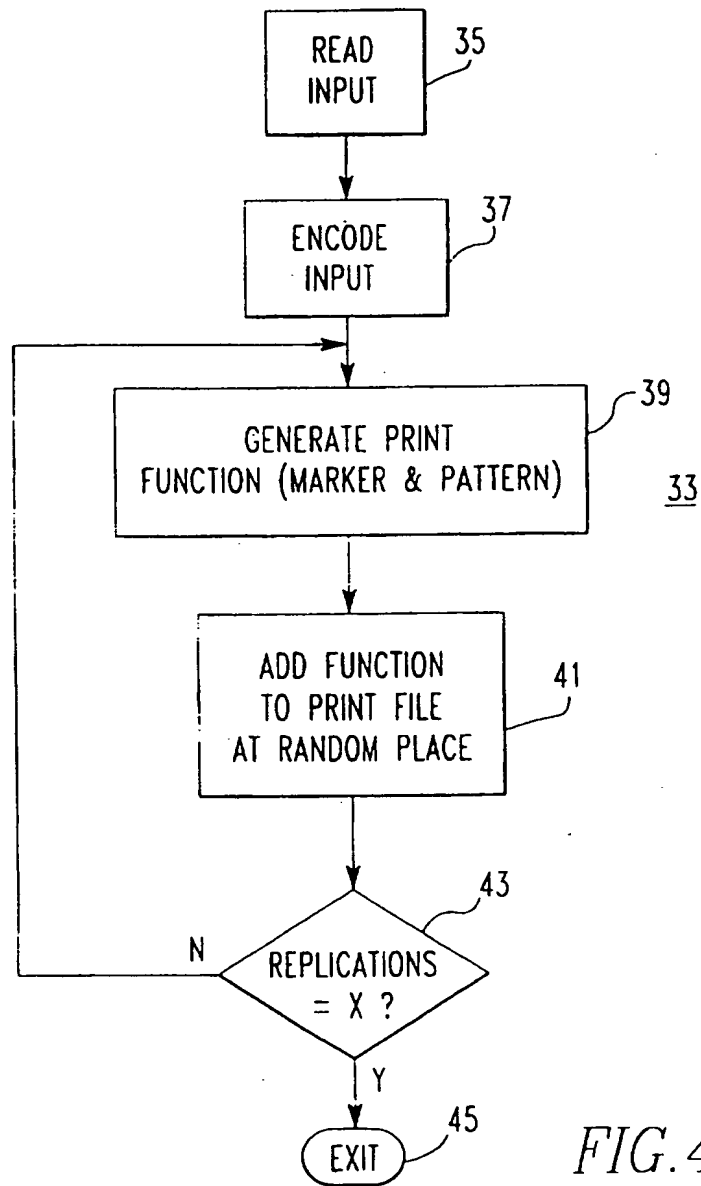FIG.2

FIG.3

READ
INPUT   35

ENCODE
INPUT   37

GENERATE PRINT
FUNCTION (MARKER & PATTERN)   39

33

ADD FUNCTION
TO PRINT FILE   41
AT RANDOM PLACE

REPLICATIONS   43
= X ?

N

Y

EXIT   45

FIG.4

A. CLASSIFICATION OF SUBJECT MATTER

IPC 6   H04N1/32

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 6   H04N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US,A,5 467 169 (MORIKAWA TAKESHI) 14 November 1995 see column 3, line 30 - column 6, line 31 ---	1-3,6, 9-13, 15-17
A	EP,A,0 551 016 (CANON KK) 14 July 1993 see column 6, line 5 - line 11 see column 11, line 33 - line 38 see column 12, line 24 - line 34 see column 13, line 46 - line 49 see column 14, line 22 - line 58 ---	1,11,15
A	PROCEEDINGS OF THE SPIE, vol. 2615, 23 October 1995, pages 226-259, XP000601271 SANDFORD M T ET AL: "THE DATA EMBEDDING METHOD" see page 240, line 12 - line 22 ---	1,11,15

-/--

X	Further documents are listed in the continuation of box C.		X	Patent family members are listed in annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
7 January 1997	30.01.97

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+ 31-70) 340-2040, Tx. 31 651 epo nl, Fax (+ 31-70) 340-3016	Hazel, J

1

page 1 of 2

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	WO,A,95 20291 (MOR LTD ;PAATELMA OTSO (GB); BORLAND ROD HUGH (GB)) 27 July 1995 ---	
A	ELECTRONICS & COMMUNICATIONS IN JAPAN, PART I - COMMUNICATIONS, vol. 74, no. 8, 1 August 1991, pages 30-36, XP000287493 KIYOSHI TANAKA ET AL: "A DIGITAL SIGNATURE SCHEME ON A DOCUMENT FOR MH FACSIMILE TRANSMISSION" -----	

1

Patent document cited in search report	Publication date	Patent family member(s)		Publication date
US-A-5467169	14-11-95	JP-A-	7023157	24-01-95
EP-A-0551016	14-07-93	CA-A-	2086472	07-07-93
		JP-A-	5301380	16-11-93
		US-A-	5457540	10-10-95
WO-A-9520291	27-07-95	AU-A-	1422895	08-08-95